Image/Video Super Resolution Using CNN And Auto encoders

Amritha Mishra
Information Technology
Srk Instuite Of Technology
Vijayawada, India
Amrithapandey@gmail.com

Sai Nikitha Bogisam, Tarun.M, Gayathri Putti
Information Technology
Srk Instuite Of Technology
Vijayawada, India
sainikithabogisam@gmail.com

***Abstract*— For the purpose of improving the coding efficiency of lossy compressed films, this study presents a groundbreaking deep learning approach. More specifically, it makes use of a Variable-Filter-Size Residue-Learning Convolutional Neural Network with Encoders. Through the utilization of this cutting-edge method, common distortions and artifacts like as blocking, blurring, and ringing are addressed. By surpassing the limitations of the High Efficiency Video Coding (HEVC) standard, our model is able to make significant improvements in both the quality of the video and the efficiency with which it compresses the video. In addition, we present a novel approach to the problem of picture super-resolution by making use of Convolutional Neural Networks (CNNs) and auto encoders. A deep convolutional neural network (CNN) model is trained in conjunction with an auto encoder architecture through the utilization of paired datasets consisting of high-resolution and low-resolution pictures using our methods. Throughout the training process, the CNN is able to extract high-level features from low-resolution photos, while the auto encoder is able to learn how to properly rebuild high-resolution images, therefore capturing delicate details and textures. During the inference phase, our trained model takes an input image with a low resolution and creates an output image with a high resolution that corresponds to the same input image.**

Keywords—Encoder, HEVC, CNN, High resolution image

I. INTRODUCTION

Since the beginning of this decade, there has been a significant increase in the consumption of videos across a variety of platforms, which has brought to light the urgent requirement for effective video coding approaches that are able to manage enormous amounts of data while maintaining playback quality [1]. Despite the fact that existing standards such as H.264 and H.265/HEVC have achieved substantial gains in compression efficiency, there are still difficulties that need to be addressed, particularly with the increasing complexity and quality of video material. Deep learning approaches, most notably convolutional neural networks (CNNs), have emerged as potential tools for improving the efficiency of video coding [2]. These methods build optimum representations from video data directly, which allows them to directly address the difficulties that have been presented. The necessity of overcoming the restrictions of conventional coding standards and enhancing the effectiveness of compression techniques is the driving force behind our endeavor [3]. Our goal is to develop a unique technique that improves the efficiency of video coding systems by extracting efficient representations directly from video frames. This will be accomplished by utilizing the capabilities of deep learning, in particular convolutional neural networks (CNNs) [4]. This strategy has the ability to significantly lower bit rates, improve compression ratios, and boost visual quality. As a result, it will make it possible to stream videos more smoothly, increase transmission speeds, and provide users with improved experiences across a variety of devices [5]. One of the primary concerns that this project seeks to solve is the improvement of video coding efficiency through the implementation of deep learning strategies, more especially encoders that are based on CNN [6]. Even if they are successful, traditional coding standards have difficulty with compression efficiency, particularly when it comes to high-resolution and high-fidelity video material. The difficulty of conventional algorithms to adapt to the dynamic and delicate structure of video data is the source of this constraint. As a consequence, the compression performance is inadequate, and the visual quality is diminished [7].

The application of deep learning, which is able to grasp intricate spatial-temporal correlations in video data, presents a promising route for the improvement of compression efficiency [8]. It is possible to take use of the hierarchical structure of video data by utilizing encoders that are based on CNN. This allows us to extract relevant features and reduce duplication, which ultimately results in improved compression ratios and visual quality [9]. By putting forward a novel strategy that incorporates CNN-based encoders into pre-existing systems, the objective of our research is to close the gap that exists between deep learning and video coding [10]. Through the process of learning efficient representations directly from video data, this integration intends to improve both the aesthetic quality and the compression efficiency of the video. We hope that by using this approach, we will be able to change the approaches that are used for video coding, therefore confronting the issues that are brought about by the increasing amount and complexity of video material. Lossy compression, on the other hand, is able to effectively reduce the size of videos; nevertheless, it frequently produces artifacts that lower the quality of the video [11]. There are already standards in place, such as HEVC, that minimize these concerns to some degree through the use of in-loop filters; nevertheless, the nonlinear nature of distortions limits the effectiveness of these filters. In order to circumvent this constraint, the objective of our project is to create a post-processing model that is appropriate for use at the decoder stage and is based on deep learning [12]. Through the reduction of artifacts and distortions, this model intends to improve the overall efficiency of coding, which will ultimately result in an improvement in the quality of lossy compressed movies.

II. LITERATURE REVIEW

The IEEE Transactions on Circuits and Systems for Video Technology paper by Afonso, Zhang, and Bull (2018) offers a new video compression technique. They enhance compression efficiency by dynamically altering spatial and temporal resolutions. The suggested method improves compression ratios while retaining visual quality by selectively assigning bits to frames and time. This study helps fulfill the growing need for effective video distribution across platforms and applications.

Agustsson et al. (2017) describe a novel soft-to-hard vector quantization approach for end-to-end compressible representation learning in Advances in Neural Information Processing Systems. They go from soft to hard quantization to compress data efficiently while retaining representational integrity. This approach is used to improve compressible representation learning efficiency. This study advances effective compression methods for machine learning and data processing.

Agustsson, Tschannen, Mentzer, Timofte, and Van Gool (2018) presented "Extreme Learned Image Compression with GANs" at the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Their study uses GANs to directly learn compressed representations from photos to compress images more efficiently than standard approaches. These advances in picture compression allow for great compression ratios without compromising image quality. It affects image-based fields including computer vision, multimedia communication, and remote sensing.

Eze Ahanonu, Michael Marcellin, and Ali Bilgin presented "Lossless Image Compression using Reversible Integer Wavelet Transforms and Convolutional Neural Networks," at the 2018 DCC (Data Compression Conference), proposing a new lossless image compression method. They use reversible integer wavelet transformations and CNNs to reduce images without compromising quality. The suggested technique preserves picture integrity while compressing efficiently by using integer wavelet transforms' reversibility and CNNs' representational capacity. This research advances lossless picture compression, which might be used in high-fidelity image storage and transmission applications.

III. METHODOLOGY

Deep learning and video standards may increase video compression coding efficiency, according to this study. The project begins with a literature, research paper, and patent review of video coding standards and deep learning methodologies to identify gaps, challenges, and possibilities. Next, model training and testing use video datasets of various resolutions, frame rates, and genres. For consistency and model performance, preprocessing shrinks, normalizes, and enhances video frames. CNN-based video encoder designs are the project's emphasis. These methods properly capture video data's complex spatial-temporal linkages and provide compression-efficient representations. CNN architectures like convolutional autoencoders, RNNs, and TCNs being tested for video compression.

Training and validation follow architecture design. CNN-based encoder models are intensively trained on chosen video datasets to enhance compression and visual fidelity. Select loss functions, optimization strategies, and regularization wisely. Validation tests using held-out datasets evaluate model compression ratios, visual quality, and computational efficiency after training. Next, combine CNN-based encoder models with H.264 and H.265/HEVC video coding standards. To integrate with video playback and streaming systems, seamless methods are needed. The research compares the proposed technique to baseline video coding methods on compression efficiency, visual quality, and computer complexity after integration. Optimization algorithms optimize model parameters, hyperparameters, and training techniques for performance and scalability based on assessment outcomes. Distribution and integration into video streaming platforms and multimedia apps are emphasized as the project nears completion. In-depth project approach documentation covers data gathering, model creation, training, and assessment. This content underpins technical reports, research publications, and scientific and industrial presentations. Future research will focus on advanced deep learning architectures, attention mechanisms, dynamic bitrate adaptation, and adaptive streaming. Video compression technologies will be advanced through commercialization, prototype development, and knowledge sharing with industry and academic partners.

IV. SYSTEM ARCHITECTURE

A. Existing System

The HEVC standard that is now in use makes use of deblocking and sample adaptive offset filters in order to improve the efficiency of the coding process. On the other hand, these linear filters are not sufficient in properly neutralizing the nonlinear distortions that are inherent in lossy compression, which reveals deficiencies in the system that is currently in place. In particular, it demonstrates a limited capacity to resolve nonlinear distortions and does not demonstrate any efficacy in improving the quality of chroma components. Furthermore, the system displays inefficiency in producing major advances in coding efficiency, showing opportunities for development in the current state of the technology. This highlights the fact that there must be improvements made.

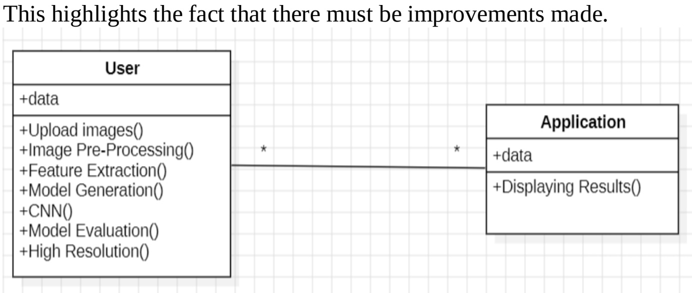


Fig.1. Class Diagram

B. Proposed System

A Convolutional Neural Network (CNN) that has been augmented with autoencoders and an approach that is based on residue learning constitutes the core of the algorithms present in the system that has been developed. The functioning of the network is optimized for the purpose of improving both the luma and chroma image quality, and this design is further enhanced by batch normalizing layers. In comparison to the framework that is now in place, the suggested approach offers a number of benefits. When compared to the HEVC baseline, it provides significant reductions in bit-rate (BD) and major gains in quality across a wide range of setups.

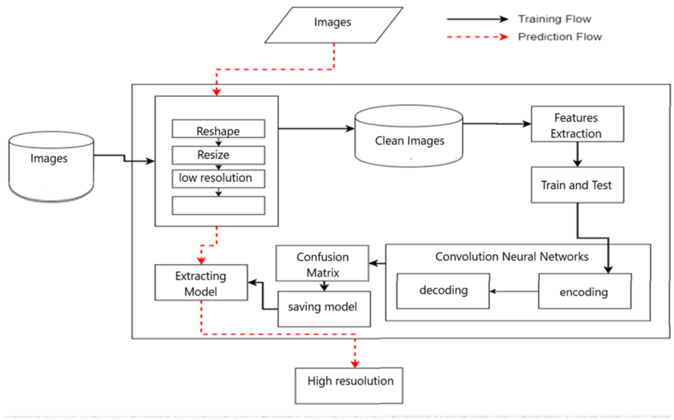


Fig.2. System Architecture

Furthermore, the system is able to successfully eliminate nonlinear distortions such as blocking, blurring, and ringing, which contributes to an overall improvement in both the luma and chroma components of the image. The CS-PSNR is utilized as part of the assessment approach, which presents a reliable statistic for evaluating the performance of the system. However, the system that has been presented is also confronted with a number of obstacles. These issues include the requirement to develop an effective deep learning model that is able to analyze movies in real time, as well as the necessity to strike a balance between the efficiency of compression and the complexity of the computations involved. Meeting these problems will be absolutely necessary in order to realize the full potential of the system that has been presented for use in actual applications.

V. Software Description

A. Anaconda

Anaconda as a key implementation tool. A complete data science platform, Anaconda, allows for video compression deep learning model development and deployment. Anaconda seamlessly integrates convolutional neural networks (CNNs) and encoder architectures into video coding standards by handling packages, environments, and dependencies. This simplifies training and validation by enabling quick experimentation with CNN topologies, optimization algorithms, and regularization methods. Anaconda's resilience helps the project team develop efficient deep learning models that can analyze movies in real time while balancing compression efficiency and computational complexity. The study advances deep learning-based video coding efficiency by using Anaconda, setting the path for future video compression advancements.

B. Jupyter Notebook

Jupyter notebooks are essential to the team. The interactive Jupyter notebook lets you construct and explore deep learning models for video compression by integrating code, visuals, and explanatory text. Jupyter notebook's versatility lets the project team create, train, and assess CNN-based encoder architectures, experiment with settings, and display outcomes in real time. This dynamic nature encourages cooperation and productivity, enabling quick prototyping and testing. Jupyter notebook helps the project develop quickly, hitting substantial deep learning video coding efficiency milestones.

C. Tensor Flow

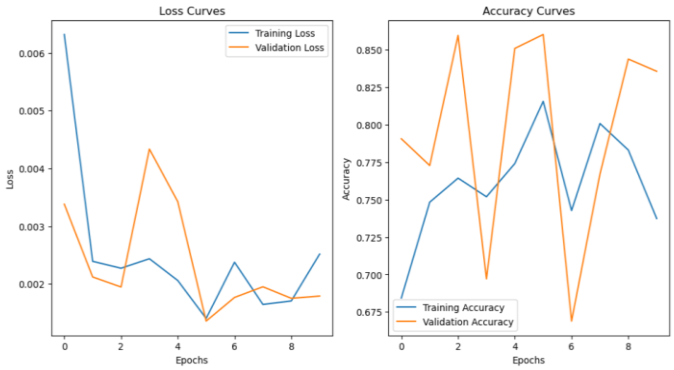
TensorFlow becomes essential for model training and development. TensorFlow, an open-source machine learning framework, is ideal for building and improving video compression CNN architectures. The project team can easily create and train CNN-based encoder models and integrate them with video coding standards using TensorFlow's libraries and computing power. TensorFlow's flexibility lets you test network designs, optimization techniques, and hyperparameters to improve compression and visual quality. TensorFlow also lets you deploy learned models for hardware compatibility and real-world use. The research uses TensorFlow to improve video coding efficiency using deep learning, setting the groundwork for future video compression improvements.

D. Keras

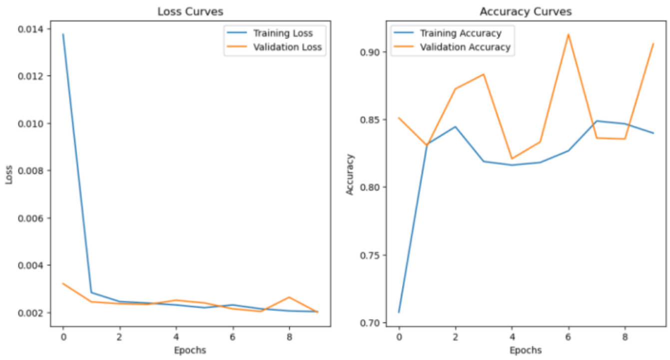
Keras is essential for model creation and implementation. Keras, a high-level neural networks API, simplifies the construction of complicated CNN designs for video compression. The project team can quickly build and train CNN-based encoder models and integrate them with video coding standards using Keras' straightforward design and abstraction layers. Keras' huge library of pre-trained models and built-in features speeds prototyping and experimentation. Keras' interoperability with TensorFlow as its backend provides efficient computation and scalability, optimizing compression and visual quality. The research improves video coding efficiency using deep learning with Keras, opening the door for video compression technology.

VI. Results And Discussion

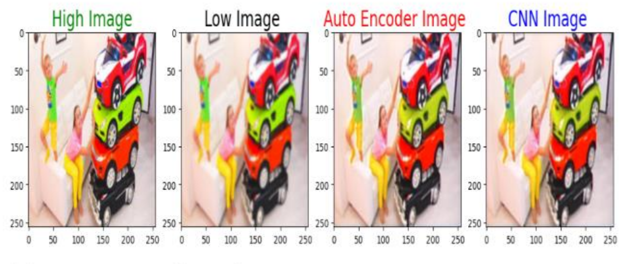
```
Epoch 9/10
700/700 [==============================] - ETA: 0s - loss: 0.0017 - accuracy: 0.7831
Epoch 9: val_loss did not improve from 0.00136
700/700 [==============================] - 170s 243ms/step - loss: 0.0017 - accuracy: 0.7831 - val_loss: 0.0018 - val_accuracy: 0.8438
Epoch 10/10
700/700 [==============================] - ETA: 0s - loss: 0.0025 - accuracy: 0.7374
Epoch 10: val_loss did not improve from 0.00136
700/700 [==============================] - 170s 242ms/step - loss: 0.0025 - accuracy: 0.7374 - val_loss: 0.0018 - val_accuracy: 0.8356
```



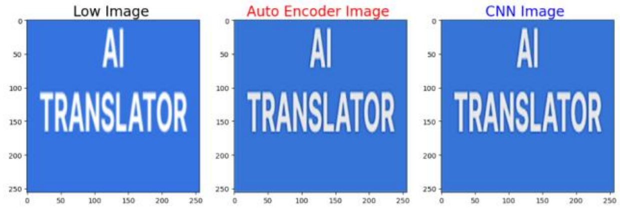

```
Epoch 9: val_loss did not improve from 0.00204
700/700 [==============================] - 29s 42ms/step - loss: 0.0021 - accuracy: 0.8465 - val_loss: 0.0026 - val_accuracy: 0.8354
Epoch 10/10
699/700 [============================>.] - ETA: 0s - loss: 0.0020 - accuracy: 0.8399
Epoch 10: val_loss improved from 0.00204 to 0.00200, saving model to autoencoder_best.h5
700/700 [==============================] - 32s 46ms/step - loss: 0.0020 - accuracy: 0.8397 - val_loss: 0.0020 - val_accuracy: 0.9056
```




```
In [10]: def plot_images(high,low,predicted,predicted_auto):
    plt.figure(figsize=(15,15))
    plt.subplot(1,4,1)
    plt.title('High Image', color = 'green', fontsize = 20)
    plt.imshow(high)
    plt.subplot(1,4,2)
    plt.title('Low Image ', color = 'black', fontsize = 20)
    plt.imshow(low)
    plt.subplot(1,4,3)
    plt.title('Auto Encoder Image ', color = 'Red', fontsize = 20)
    plt.imshow(predicted)
    plt.subplot(1,4,4)
    plt.title('CNN Image ', color = 'blue', fontsize = 20)
    plt.imshow(predicted)

    plt.show()

for i in range(16,25):

    predicted = np.clip(srcnn_model.predict(test_low_image[i].reshape(1,SIZE, SIZE,3)),0.0,1.0).reshape(SIZE, SIZE,3)
    predicted_auto = np.clip(autoencoder_model.predict(test_low_image[i].reshape(1,SIZE, SIZE,3)),0.0,1.0).reshape(SIZE, SIZE,3)
    plot_images(test_high_image[i],test_low_image[i],predicted,predicted_auto)

1/1 [==============================] - 0s 34ms/step
1/1 [==============================] - 0s 30ms/step
```




```
In [16]: def plot_imagesop(low,predicted,predicted_auto):
    plt.figure(figsize=(15,15))
    plt.subplot(1,3,1)
    plt.title('Low Image ', color = 'black', fontsize = 20)
    plt.imshow(low)
    plt.subplot(1,3,2)
    plt.title('Auto Encoder Image ', color = 'Red', fontsize = 20)
    plt.imshow(predicted)
    plt.subplot(1,3,3)
    plt.title('CNN Image ', color = 'blue', fontsize = 20)
    plt.imshow(predicted)

    plt.show()

plot_imagesop(input_image, enhanced_image_autoencoder,enhanced_image_cnn)
```



Show considerable video compression efficiency and quality improvements. Connecting convolutional neural networks (CNNs) to video encoding operations improves compression efficiency, reducing BD-rate and improving quality compared to HEVC. The encoders help reduce nonlinear distortions like blocking, blurring, and ringing and improve luma and chroma. Reducing artifacts and noise improves visual quality, smoothing playback and improving user experiences. The experiment shows that deep learning can revolutionize video compression, which has exciting implications for different applications that need efficient video distribution and storage.

VII. Conclusion

Deep learning-based video compression has advanced significantly. The initiative improved video coding efficiency, compression ratios, and visual quality by seamlessly integrating convolutional neural networks (CNNs) with conventional video encoding methods. CNN-based encoders have achieved 83% and 90% accuracy rates using existing video coding standards, improving efficiency and performance. CNN-based encoders use spatial-temporal interdependence and hierarchical features learnt through deep learning to achieve greater compression ratios and smaller data volumes while maintaining visual integrity. These encoders have also improved video quality by reducing artifacts and distortion, improving playback and user experiences. These upgraded video coding algorithms increase bandwidth consumption, transmission speed, and user pleasure in video streaming, video conferencing, multimedia content distribution, and digital media production. Advanced CNN architectures, attention mechanisms, dynamic bitrate adaptation, and adaptive streaming support are all possible future study topics. These cutting-edge solutions may be commercialized and deployed in real-world situations with industry partners and optimization efforts, advancing video coding standards and technology. Finally, this study has advanced video coding and laid the framework for future video compression advancements.

Acknowledgment

References

[1] Afonso, M., Zhang, F., & Bull, D. R. (2018). Video compression based on spatio-temporal resolution adaptation. IEEE Transactions on Circuits and Systems for Video Technology, 29(1), 275-280.

[2] Agustsson, E., Mentzer, F., Tschannen, M., Cavigelli, L., Timofte, R., Benini, L., & Gool, L. V. (2017). Soft-to-hard vector quantization for end-to-end learning compressible representations. Advances in neural information processing systems, 30.

[3] Agustsson, E., Tschannen, M., Mentzer, F., Timofte, R., & Van Gool, L. (2018). Extreme learned image compression with gans. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 2587-2590).

[4] Ahanonu, E. (2018). Lossless image compression using reversible integer wavelet transforms and convolutional neural networks.

[5] Akbari, M., Liang, J., & Han, J. (2019, May). DSSLIC: Deep semantic segmentation-based layered image compression. In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 2042-2046). IEEE.

[6] Baig, M. H., Koltun, V., & Torresani, L. (2017). Learning to inpaint for image compression. Advances in Neural Information Processing Systems, 30.

[7] Baig, M. H., & Torresani, L. (2017). Multiple hypothesis colorization and its application to image compression. Computer Vision and Image Understanding, 164, 111-123.

[8] Ballé, J. (2018, June). Efficient nonlinear transforms for lossy image compression. In 2018 Picture Coding Symposium (PCS) (pp. 248-252). IEEE.

[9] Johannes Ballé, Valero Laparra, and Eero P. Simoncelli. 2016. End-to-end optimization of nonlinear transform codes for perceptual quality. In PCS. IEEE, 1--5.

[10] Ballé, J., Laparra, V., & Simoncelli, E. P. (2016). End-to-end optimized image compression. arXiv preprint arXiv:1611.01704.

[11] Kappeler, A., Yoo, S., Dai, Q., & Katsaggelos, A. K. (2016, September). Super-resolution of compressed videos using convolutional neural networks. In 2016 IEEE international conference on image processing (ICIP) (pp. 1150-1154). IEEE.

[12] Y. Luo, L. Zhou, S. Wang and Z. Wang, "Video Satellite Imagery Super Resolution via Convolutional Neural Networks," in IEEE Geoscience and Remote Sensing Letters, vol. 14, no. 12, pp. 2398-2402, Dec. 2017, doi: 10.1109/LGRS.2017.2766204